RESTORATION OF DEPTH AND INTENSITY IMAGES USING A GRAPH LAPLACIAN REGULARIZATION

Abderrahim Halimi(1), Peter Connolly(1), Ximing Ren(1), Yoann Altmann(1), Istvan Gyongy(2),
Robert. K. Henderson(2), Stephen McLaughlin(1), Gerald S. Buller(1)

(1) Heriot-Watt University, School of Engineering and Physical Sciences, Edinburgh U.K.

ABSTRACT

This paper presents a new algorithm for the joint restoration of depth and intensity (DI) images constructed using a gated SPAD-array imaging system. The three dimensional (3D) data consists of two spatial dimensions and one temporal dimension, and contains photon counts (i.e., histograms). The algorithm is based on two steps: (i) construction of a graph connecting patches of pixels with similar temporal responses, and (ii) estimation of the DI values for pixels belonging to homogeneous spatial classes. The first step is achieved by building a graph representation of the 3D data, while giving a special attention to the computational complexity of the algorithm. The second step is achieved using a Fisher scoring gradient descent algorithm while accounting for the data statistics and the Laplacian regularization term. Results on laboratory data show the benefit of the proposed strategy that improves the quality of the estimated DI images.

Index Terms— Graph Laplacian regularization, Bayesian inference, Image restoration, Fisher scoring.

1. INTRODUCTION

In addition to intensity levels, three dimensional (3D) imaging systems provide information regarding the target’s depth [1,2]. Single-photon systems present clear advantages thanks to their photon sensitivity and good performance under low illumination imaging conditions [3]. In this paper, a single-photon depth profiling system is used to acquire 3D images by illuminating the scene with a train of laser pulses and capturing the reflected photons using a gated CMOS single photon avalanche diode (SPAD) camera [4, 5]. The latter contained an array of 256 × 256 pixels and is time gated. To explore the range dimension (i.e., depth), the temporal gate (of the order of nanoseconds) of the camera is time shifted by considering small steps (of the order of picoseconds) to ensure a good depth resolution. At each temporal delay, a set of \( N_f \) binary images is acquired and summed to obtain an image of photon counts. This leads to 3D data, which contains photon counts, and corresponds to two spatial dimensions representing pixels and one temporal dimension related to the range direction. In absence of a target inside the system gate, the camera only captures background noise (see delays 1 to 100 in Fig. 1 (b)). However, when the target is inside the system gate, the temporal response presents an increase with respect to (w.r.t.) delays whose amplitude, slope and position are related to the target’s reflectivity, curvature/roughness and depth, respectively. In this paper, we are interested in the extraction of the 3D information provided by these data together with the restoration of the estimated images.

To improve the quality of the estimated images, the first contribution of this paper is to consider a non-local approach [6, 7] by using a graph Laplacian regularization [8, 9]. Assuming pixels with similar temporal response to provide similar estimated images, a graph can be built by considering the pixels as its nodes and their similarities as its edges. However, such a graph will have a large dimension leading to computational difficulties. Therefore, we propose in this paper to generalize the approach proposed in [9] by dividing the large graph into smaller sub-graphs. A multi-level approach is considered, where the graph-nodes are associated with patches of pixels that have the same size for each level. Then, a clustering is performed on the resulting graph leading to smaller sub-graphs, that are computationally more attractive.

A second contribution is related to the parameter estimation that is achieved by combining a hierarchical Bayesian model with a natural gradient descent algorithm. The hierarchical Bayesian model accounts for the data binomial statistics, the non-linear observation model related to the parameters, and the non-local regularization term. This model introduces hyperparameters that are marginalized to avoid their estimation. The maximum-a-posteriori (MAP) estimate associated with the resulting model is approximated using a natural gradient algorithm that is known for its fast convergence properties [10, 11]. The proposed approach is validated on laboratory data showing improved estimates for the different parameters.

This paper is structured as follows. Section 2 presents
the observation model and provides details regarding the construction of the Laplacian graph. Section 3 describes the proposed hierarchical Bayesian model. The parameters associated with the proposed model are estimated in Section 4. Section 5 shows simulation results obtained on laboratory data. Conclusions and future work are finally reported in Section 6.

2. OBSERVATION MODEL AND GRAPH MAPPING

This paper considers a gated imaging system, which provides both reflectivity and depth information regarding the observed target. This system provides photon counts that are gathered in a three-dimensional data cube of two spatial dimensions and one dimension related to the range. Let \( y_{n,t} \) where \( n \in \{1, \cdots, N\} \) and \( t \in \{1, \cdots, T\} \), denotes the number of photon counts within the \( t \)th delay of the \( n \)th pixel. This variable \( y_{n,t} \) contains less than \( N_f \) photons and is assumed to be distributed according to a binomial distribution \( B(.) \) as follows

\[
y_{n,t} \sim B \left( N_f, \frac{s_{n,t}}{N_f} \right)
\]

where \( s_{n,t} \) denotes the average photon counts whose shape is related to the system impulse response. In this paper, we approximate it by the following non-linear model

\[
s_{n,t} = \frac{r_n}{2} \left( 1 + \text{erf} \left( \frac{t - d_n}{h_n} \right) \right) + b_n
\]

where \( \text{erf}(.) \) denotes the error function, \( d_n \geq 0 \) is the range of the target (related to its depth), \( r_n \geq 0 \) is related to the target’s intensity, \( b_n \geq 0 \) denotes the background and dark counts of the detector, \( h_n \geq 0 \) is related to the curvature/roughness of the observed target’s surface and \( \Theta = (\theta_1, \cdots, \theta_4) = (r, d, h, b) \) is the \( (N \times 4) \) matrix gathering all the parameters of interest. In addition to the good approximation of model (2) to the measured pixels (see Fig. 1 (b)), its parameters have a direct physical interpretation as the target’s depth and intensity that are related to the function’s temporal shift and amplitude, respectively. The proposed method aims at estimating the matrix of parameters \( \Theta \) while considering the statistics of the data in (1) and spatially non-local smoothness constraints on the estimated images.

2.1. Non-local spatial correlation

Non-local regularization has received an increased interest in the image processing community thanks to its good performance [6, 7]. This paper introduces a non-local smoothness constraints on the estimated parameters by using a graph representation of the data. Consider first that each pixel spectrum is associated to a graph node and let \( \mathbf{W} \) be the \( N \times N \) affinity matrix associated with the graph. Matrix \( \mathbf{W} \) has positive elements \( w_{i,j} \geq 0 \) that represent the degree of similarity between the nodes \( i \) and \( j \), respectively. Different heuristics can be considered for the choice of \( w_{i,j} \) such as the Gaussian kernel \( w_{i,j} = \exp \left[ -\|s_{i,t} - s_{j,t}\|^2/(2\sigma^2) \right] \), where \( \sigma \) denotes the kernel’s bandwidth [9, 12]. A Laplacian can then be computed from the affinity matrix \( \mathbf{W} \) and used to construct a regularization term to improve the quality of the estimated parameters \( \Theta \). However, the direct application of this Laplacian matrix raises some difficulties related to its large size (e.g., a 128 \( \times \) 128 image leads to a huge matrix \( \mathbf{W} \) of size \( 16384 \times 16384 \)). In this paper, we generalize the procedure proposed in [9] to achieve better computational performance. To reduce the graph size, a multilevel approach is considered. In the first level \( \ell = 1 \), the graph-node gathers a set of \( n_r^{\ell} \times n_r^{\ell} \) pixels and an affinity matrix of a reduced size (i.e., \( \frac{N}{n_r^{\ell}} \times \frac{N}{n_r^{\ell}} \)) is evaluated. Akin to [9], algorithm [13] is then applied to the resulting graph to partition its nodes into \( k_{\ell} \) subgraphs or clusters. At this step, the pixels of each sub-graph can be processed independently, which is of great interest to reduce the computational cost. In addition, the affinity matrix of each sub-graph is a subset of the original affinity graph, therefore it does not need to be recomputed. For the next level \( \ell + 1 \), the procedure is repeated by considering nodes of size \( n_r^{\ell+1} \times n_r^{\ell+1} \) pixels and a partition into \( k_{\ell+1} \) clusters, where of course \( n_r^{\ell+1} < n_r^{\ell} \). After the evaluation of \( L \) levels, the procedure provides an image partition into \( K = k_1 k_2 \cdots k_L \) clusters having their own small sub-graphs \( \mathbf{W}(k) \). Thus, the pixels of each sub-graph can be processed independently by the estimation procedure described in Section 4 leading to a reduced computational cost.

3. HIERARCHICAL BAYESIAN MODEL

The quality of the estimated images \((r, d, h, b)\) depends on the number of considered frames \( N_f \), where the higher the better. This section introduces a hierarchical Bayesian model to improve these images by accounting for the data statistics and the available a-priori knowledge about the estimates.

3.1. Likelihood

Considering the data statistical model in (1) and assuming independence between the observations yield

\[
f(Y|\Theta) \propto \prod_{t=1}^{T} \prod_{n=1}^{N} \left( \frac{s_{n,t}}{N_f} \right)^{y_{n,t}} \left( 1 - \frac{s_{n,t}}{N_f} \right)^{N_f - y_{n,t}}
\]

where \( \propto \) means “proportional to”, \( Y \) is the \( N \times T \) matrix gathering the elements \( y_{n,t}, \forall t, n \), and \( s_{n,t}(\Theta_n) \) has been denoted by \( s_{n,t} \) for brevity.

3.2. Prior distributions for \( \Theta \)

To restore the estimated depth \( d \), intensity \( r \) and slope \( h \) images, we take advantage of the spatial correlation available in these images. To do so, we encourage the \( h \)th parameter \( \theta_{h,k} \) associated with the pixels of spatial class \( k \) (determined in Section 2.1) to share similar values. More precisely, we assign a Gaussian prior distribution to \( \theta_{h,k} \) for \( i \in \{1, 2, 3\} \) and \( k \in \{1, \cdots, K\} \) as follows [9, 11]
where $L_k$ denotes the graph Laplacian operator of the $k$th class given by $L_k = G(k) - W(k)$, with $G(k)$ a diagonal matrix such that $G_{a,n}(k) = \sum_{j=1}^N w_{a,j}(k)$, and $\epsilon^2_{i,k}$ is a hyperparameter that controls the degree of promoted smoothness.

For each parameter $\theta_{i,k}$, prior (4) penalizes the square of the mutual difference between its element weighted by their degree of similarity introduced by $L$. The interested reader is invited consult [8, 9] for more details regarding this prior. Considering a similar Gaussian prior, the background noise’s level, associated with each class $k$, has been constrained as follows

$$f(b_k|\epsilon^2_{i,k}) \propto \left(\frac{1}{\epsilon^2_{i,k}}\right)^{N/2} \exp\left(-\frac{1}{2\epsilon^2_{i,k}} b_i^T L_k b_i\right).$$

### 3.3. Hyperparameter priors

Conjugate inverse gamma distributions have been assigned to the hyperparameters $\epsilon^2_{i,k}$ via $i, k$. To make the estimation procedure more robust and avoid the estimation of these hyperparameters, we propose to marginalize the joint posterior distribution $f(\Theta, \epsilon|Y)$ w.r.t. $\epsilon$ as follows

$$f(\Theta|Y) \propto f(Y|\Theta) \prod_{i,k} f(\theta_{i,k}|\epsilon^2_{i,k}) f(\epsilon^2_{i,k}|\alpha_{i,k}, \beta_{i,k}) \prod_{i,k} \right) \propto f(Y|\Theta) \prod_{i,k} f(\theta_{i,k}|\alpha_{i,k}, \beta_{i,k})$$

where we have assumed a priori independence between the parameters and hyperparameters and $f(\theta_{i,k}|\alpha_{i,k}, \beta_{i,k}) \propto \left(\frac{\theta^T L_{i,k} \theta_{i,k}}{2} + \beta_{i,k}\right)^{-\alpha_{i,k} - m_{i}/2}$ with $m_i = N - 1$ and $L_{i,k} = L_k$ for $i \in \{1, 2, 3\}$, and $m_4 = N$ and $L_{4,k}$ is the identity matrix. The next section introduces the optimization algorithm used to approximate the MAP estimator of $\Theta$ by minimizing the negative log-posterior $C = -\log[f(\Theta|Y)]$ given by

$$C(\Theta) = \sum_{t,n} \left[-y_{n,t} \log \frac{s_{n,t}}{N_f} - (N_f - y_{n,t}) \log \left(1 - \frac{s_{n,t}}{N_f}\right)\right] + \sum_{i,k} \left(\alpha_{i,k} + \frac{m_{i}}{2}\right) \log \left(\frac{\theta^T L_{i,k} \theta_{i,k}}{2} + \beta_{i,k}\right).$$

### 4. NATURAL GRADIENT ALGORITHM

The marginalized posterior (6) is maximized w.r.t. $\Theta$ using a projected natural gradient descent algorithm. Thanks to the independence between the pixels of different spatial classes, the parameter estimation can be achieved on each cluster independently. Thus, we drop the indices $k$ in this section for brevity.

#### 4.1. Natural gradient for parameter estimation

A projected natural gradient descent algorithm is used to ensure a fast estimation of the parameters [10, 11]. This algorithm updates the $(4N \times 1)$ vector of parameters $\gamma = (\theta_1^T, \ldots, \theta_4^T)^T$ as follows

$$\gamma^{(t+1)} = P_{R^+}\left[\gamma^{(t)} - F^{-1}\left(\gamma^{(t)}\right) \nabla C\left(\gamma^{(t)}\right)\right]$$

where $P_{R^+}(.)$ is a projection operator imposing positive values for the parameters (e.g., $r_n \geq 0$), the gradient

$$\nabla C\left(\gamma^{(t)}\right) = \left[\frac{\partial C}{\partial \theta_1^T}, \ldots, \frac{\partial C}{\partial \theta_4^T}\right]^T$$

and the Fisher information matrix $F(\gamma^{(t)})$ are given by

$$F = \begin{bmatrix} F_1 & F_{12} & F_{13} & F_{14} \\ F_{12} & F_2 & F_{23} & F_{24} \\ F_{13} & F_{23} & F_3 & F_{34} \\ F_{14} & F_{24} & F_{34} & F_4 \end{bmatrix}$$

with

$$F_i = F_{ii} + \left(\alpha_i + \frac{m_i}{2}\right) \left[\frac{\theta_i^T L_i \theta_i}{2} + \beta_i\right] + \left[\frac{\theta_i^T L_i \theta_i}{2} + \beta_i\right]$$

for $i \in \{1, 2, 3, 4\}$, where $F_{ij} = \text{diag} \left(\frac{x_{1,ij}}{\gamma_{1,ij}}, \ldots, \frac{x_{N,ij}}{\gamma_{N,ij}}\right)$, $\gamma = (\gamma_1, \ldots, \gamma_N)^T$, $x_{n,ij} = \sum_{t=1}^T \left[\frac{N_f}{\gamma_{n,f}} \frac{\partial c_n}{\partial \gamma_{n,f}} \frac{\partial \gamma_{n,f}}{\partial y_{n,t}}\right]$, and $z_{n,ij} = \sum_{t=1}^T \left[\frac{N_f}{\gamma_{n,f}} \frac{\partial c_n}{\partial \gamma_{n,f}} \frac{\partial \gamma_{n,f}}{\partial y_{n,t}}\right]$. In contrast to a steepest descent algorithm, (8) requires the additional computation of the inverse of $F$. The latter, however, allows a faster convergence of the algorithm as already reported in [10, 11].

#### 4.2. Stopping criteria and parameters initialization

This paper considers three stopping criteria including a maximum number of iterations $T_{\max}$ and [11, 15]

$$(\gamma^{(t+1)} - \gamma^{(t)})_2 \leq \xi_2 \left(\|\gamma^{(t)}\|_2 + \xi_2\right)$$

where $\xi_1, \xi_2$ are fixed thresholds, and $\|\cdot\|_2$ (resp. $\|\cdot\|$) denotes the absolute value (resp. norm-2).
Note that the cost function is not convex, thus, the parameters have been well initialized by considering the results of a classical method. The latter estimates the depth by performing a cross-correlation of each pixel with its impulse response (acquired during the system calibration), and the intensity by averaging the counts of a set of delays located in the plateau of each pixel (see Fig. 1 (b)). The background is also initialized by considering the average of the first 10 delays. With this initialisation, the algorithm always converges to good results as will be shown in the next section.

5. RESULTS

The performance of the proposed algorithm is evaluated on a 256 × 256 image of a life-sized polystyrene head (see Fig. 1 (a)), located at a distance of 2 m from the system. The data were acquired in May 2017 inside a laboratory of Heriot-Watt University. The system employed a pulsed illumination and a gated CMOS single photon avalanche diode (SPAD) camera. To obtain a depth profile of the target, the system’s gate (of duration 10 ns) is step-by-step shifted with a step of 10 ps using a programmable delay generator. In this paper, we consider images obtained with $N_f \in \{50, 200, 500, 1000\}$ frames, where high $N_f$ values lead to a better image quality. These data were downsampled to have 128 × 128 pixels and were processed using the proposed algorithm with 2 levels, $k_1 = 2$, $k_2 = 16$, $n^1_r = n^1_c = 8$ and $n^2_r = n^2_c = 1$. Fig. 1 (c,d) show an example of the estimated DI images for $N_f = 500$ frames while considering the proposed algorithm and the classical algorithm (i.e., the depth is estimated using a cross-correlation between each pixel and its impulse response, and the intensity by averaging the counts in the plateau of each pixel). This figure shows the benefit of the proposed algorithm in removing the noise affecting the measures. In addition to the DI images, the proposed algorithm also provides the slope parameter $h_n, \forall n$ that is related to the curvature of the target as shown in Fig. 1 (c), where a higher value means a higher curvature. Finally, Table 1 confirms the good results of the proposed algorithm when considering the signal-to-reconstruction error (SRE) ratio $\text{SRE} = 10 \log_{10} \left( \frac{||x||^2}{||x - \hat{x}||^2} \right)$, where $x$ is the reference depth or intensity image obtained by processing DI of $N_f = 1000$ frames using the proposed algorithm, and $\hat{x}$ is the estimated image.

6. CONCLUSIONS

This paper presented a new algorithm for the restoration of depth and intensity images acquired using a gated SPAD-array imaging system. An analytical model was first proposed to approximate the system impulse response. This model was then fitted to the data while accounting for their binomial statistics, and the non-local spatial similarity presents in the depth, intensity and slope images. The proposed approach showed good restoration results when processing images of a polystyrene head acquired in laboratory. Future work includes the generalization of the proposed approach to deal with scenarios showing a reduced number of frames.

### Table 1. SRE for DI images using the proposed and classical algorithms.

<table>
<thead>
<tr>
<th></th>
<th>1000</th>
<th>500</th>
<th>200</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed algorithm</td>
<td>Depth</td>
<td>Inf</td>
<td>36.1</td>
<td>34.4</td>
</tr>
<tr>
<td>Classic algorithm</td>
<td>Depth</td>
<td>Inf</td>
<td>22.2</td>
<td>22.2</td>
</tr>
<tr>
<td>Intens.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed algorithm</td>
<td>Intens.</td>
<td>18.2</td>
<td>18.2</td>
<td>18.1</td>
</tr>
<tr>
<td>Classic algorithm</td>
<td>Intens.</td>
<td>24.1</td>
<td>24.1</td>
<td>24.4</td>
</tr>
</tbody>
</table>
7. REFERENCES


